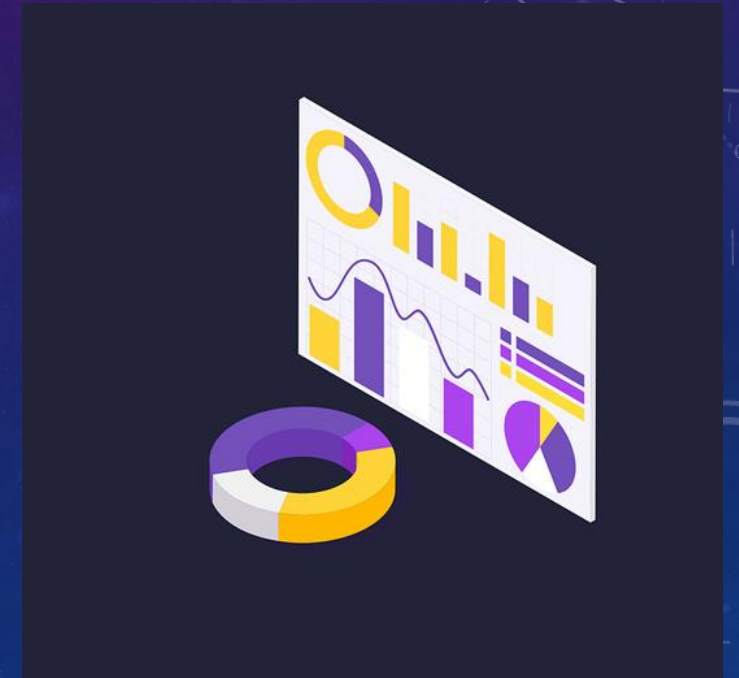


The background is a gradient of deep blue and purple, speckled with white dots resembling stars. Overlaid on this are several faint, white circular and semi-circular patterns. Some of these patterns include tick marks and numbers, suggesting a circular scale or a compass. There are also curved arrows indicating a direction of movement or flow. The overall aesthetic is technical and futuristic.

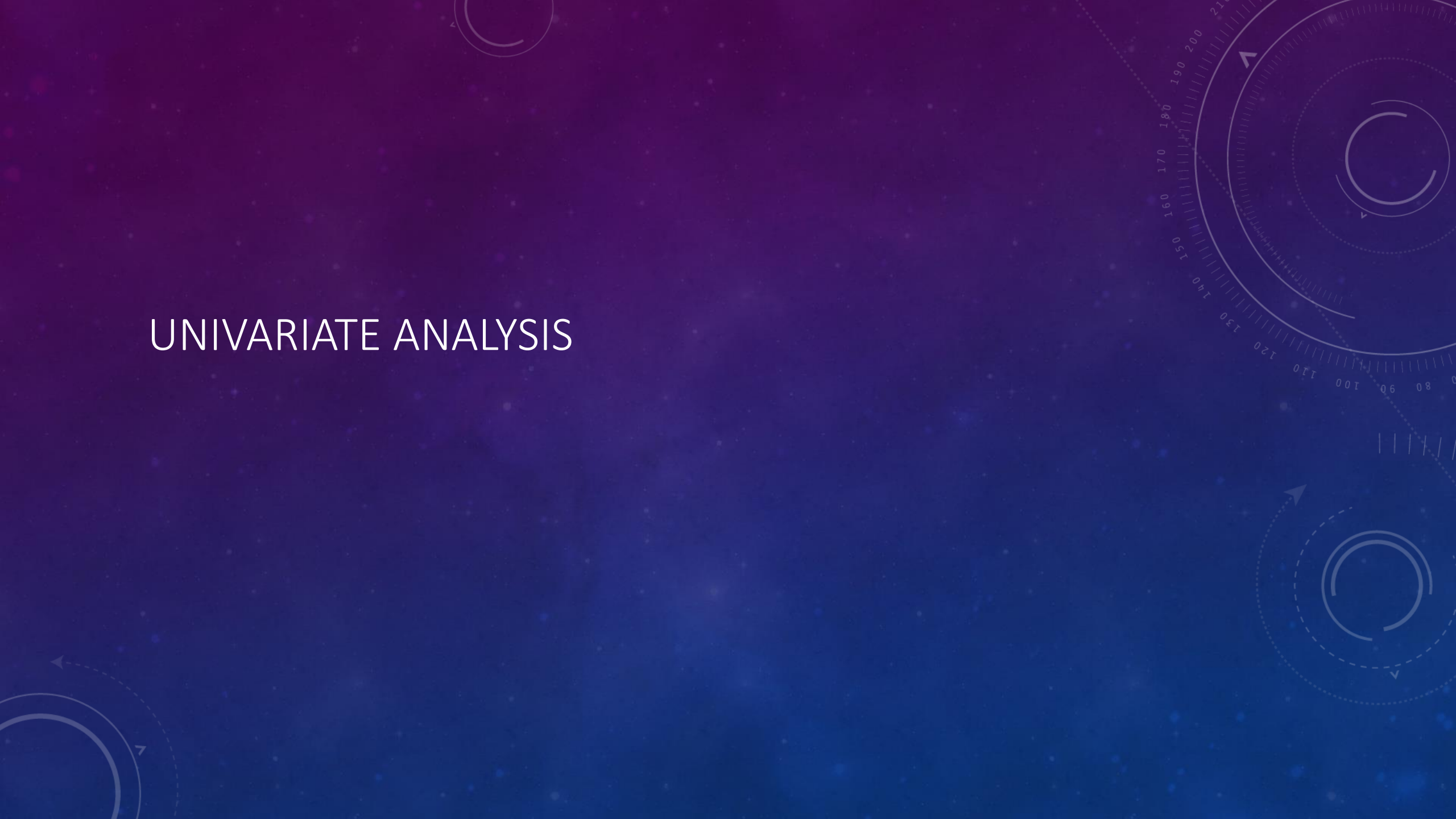
# CREDIT EDA CASE STUDY

# BUSINESS PROBLEM AND END RESULTS

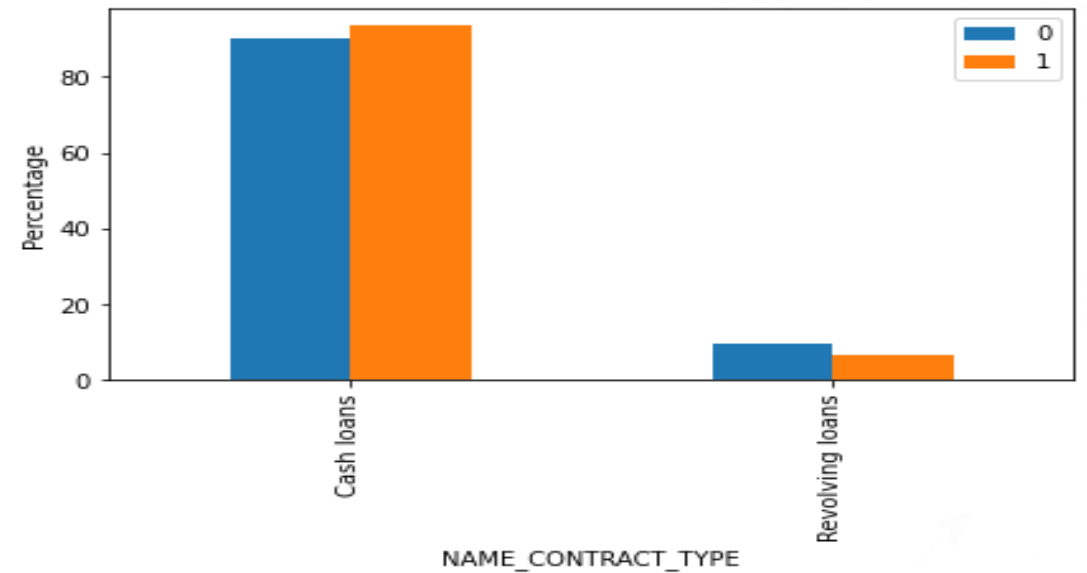
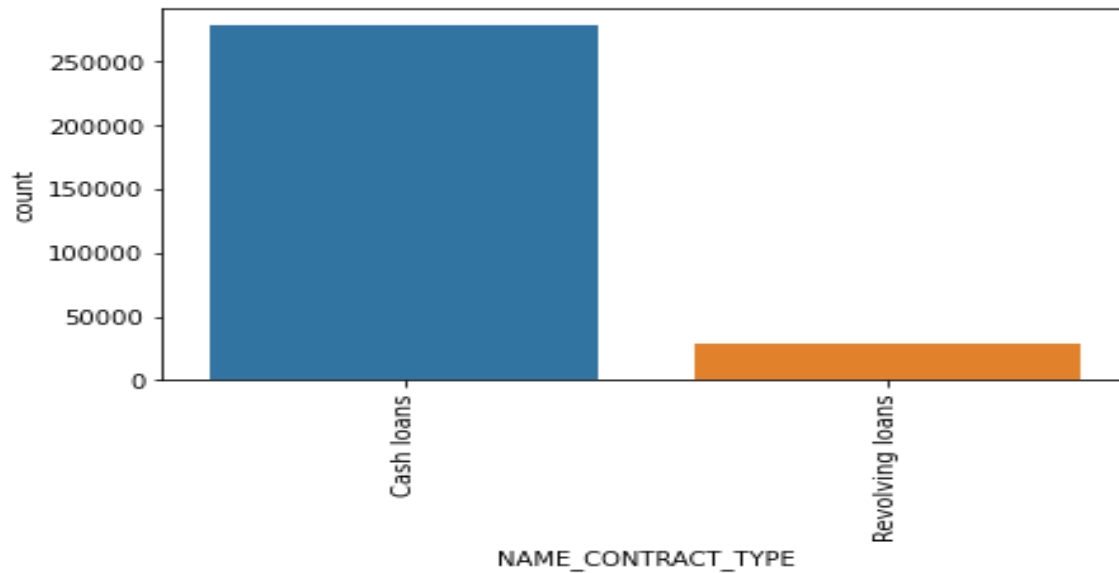
- Applying EDA in a real business scenario
- Basic understanding of risk analytics in banking and financial services.
- Understand how data is used to minimize the risk of losing money while lending to customers.
- This case study tries to uncover patterns that suggest whether a client is having trouble paying their installments, which can be used to make decisions like refusing the loan, reducing the loan amount, lending (to risky applicants) at a higher interest rate, and so on. This will ensure that those customers who are capable of repaying the loan are not turned down.



# UNIVARIATE ANALYSIS



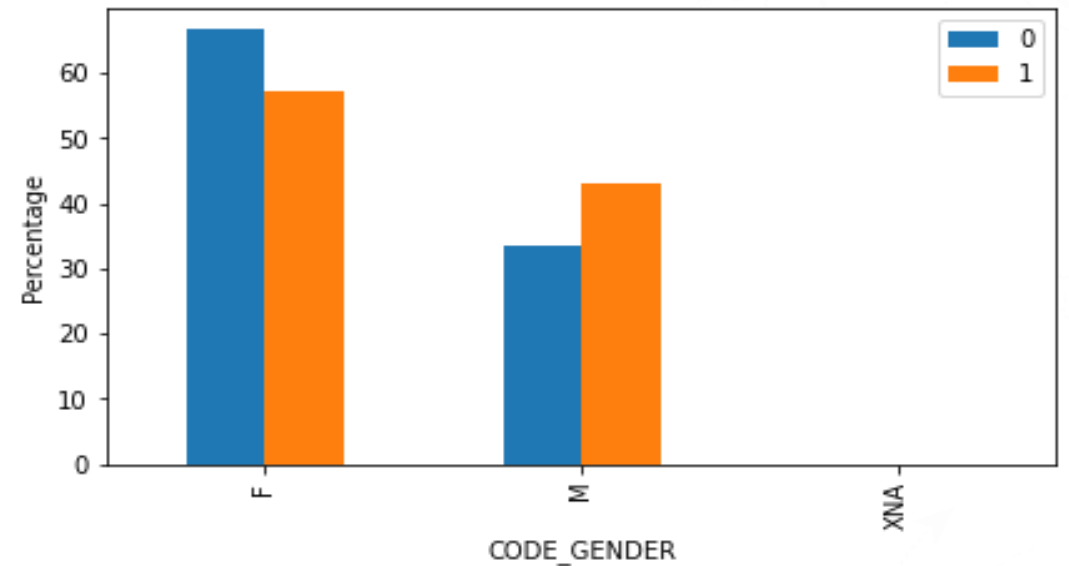
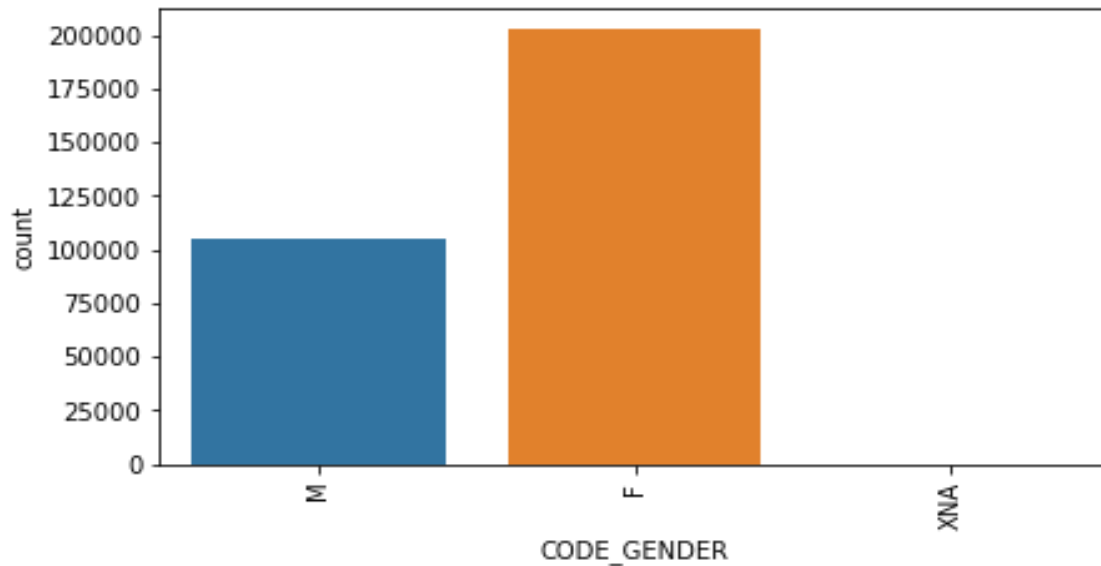
# DISTRIBUTION OF CONTRACT TYPE



From the above plots, we can state that:

- Number of loans applied are more for cash loans than revolving loans.
- Number of people who have difficulties in payment are slightly greater in cash loans as compared to others, whereas for revolving loans, number of people having difficulties in payment are slightly less than the other.

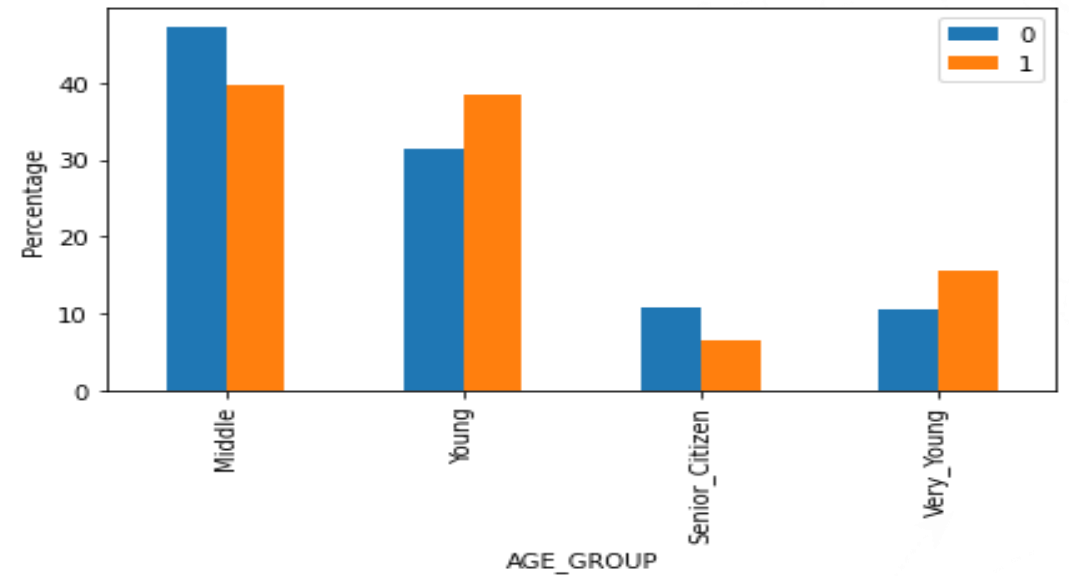
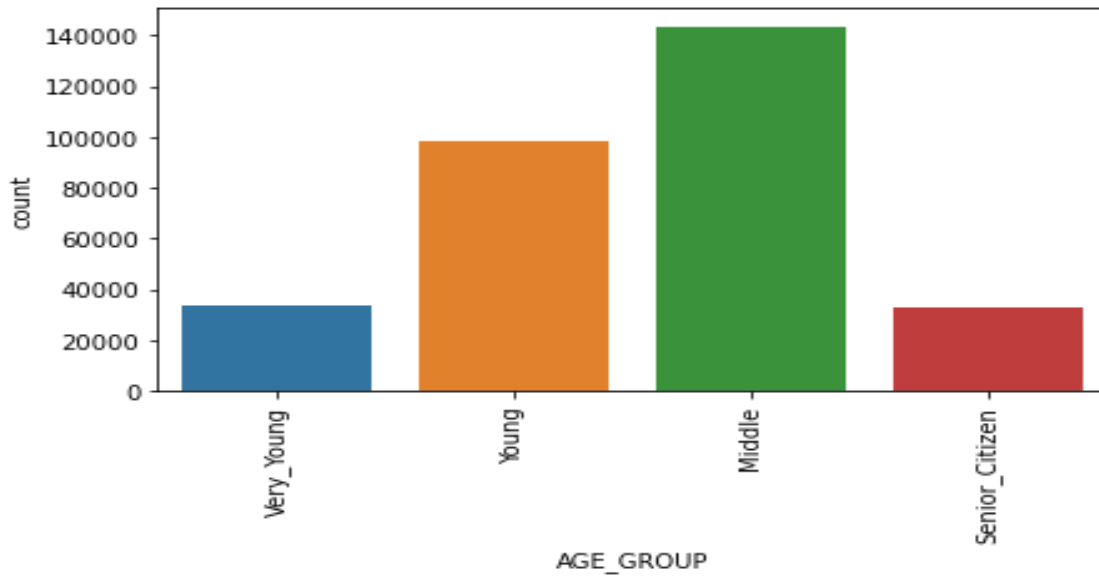
# DISTRIBUTION OF GENDER



From the above plots, we can state that:

- Females are applying for more number of loans.
- Also less females have difficulty in payment, more females have no difficulty in payment.
- For males, percentage of males having difficulty in payments is more than the percentage of males having no difficulty in payment.
- On comparing both the genders, percentage of females having difficulties in payment is more than that of males.

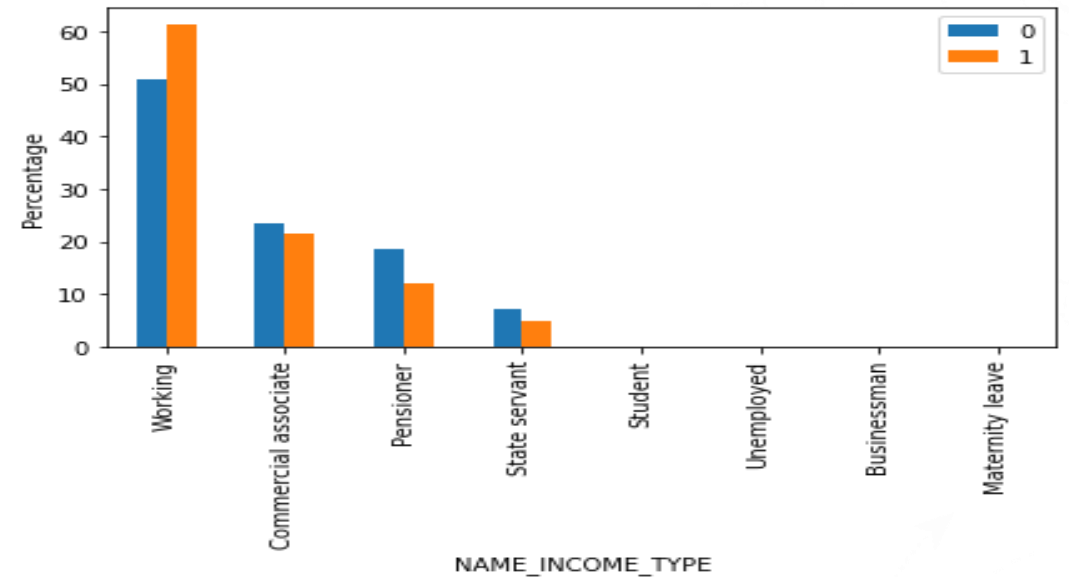
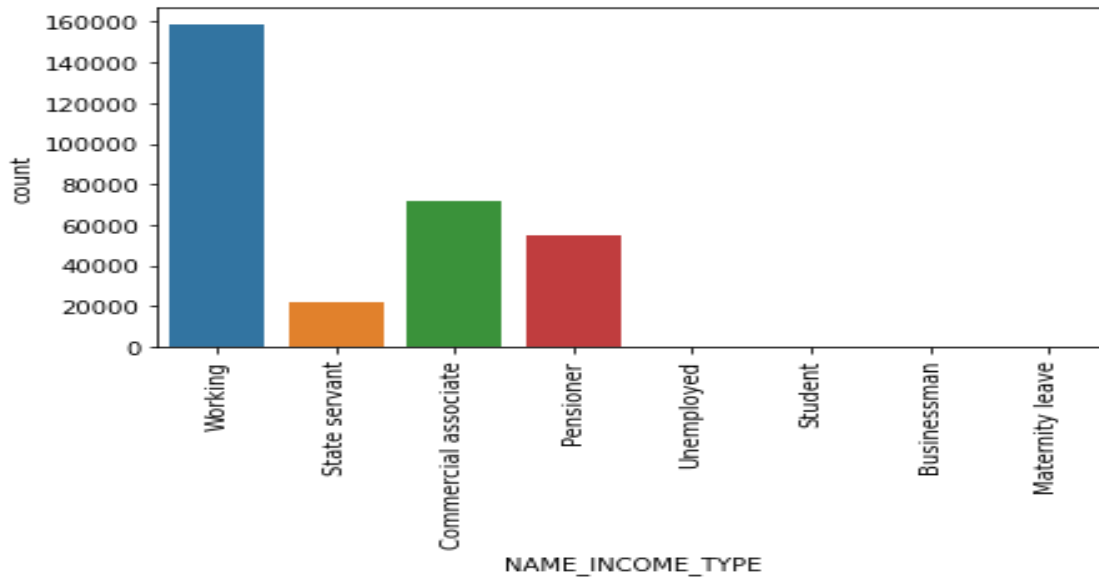
# DISTRIBUTION OF AGE



From the above plots, we can state that:

- Middle age group is applying for most number of loans.
- Very young and young age group clients have higher percentage for payment difficulties, whereas middle and senior citizen have less percentage for clients having difficulties in payment

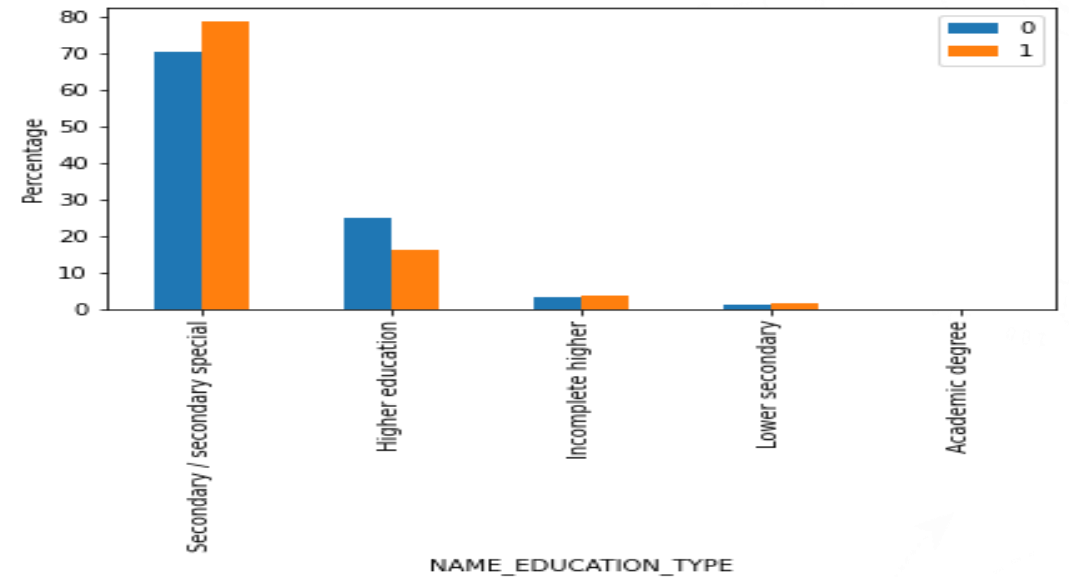
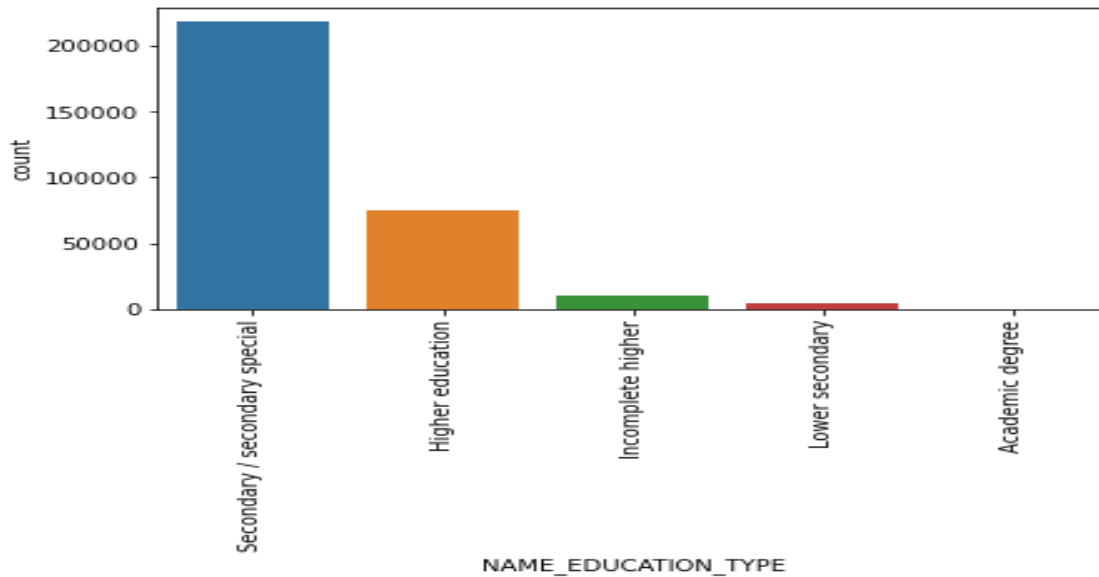
# DISTRIBUTION OF INCOME TYPE



From the above plots, we can state that:

- Out of all the categories, Working category applies for maximum loan.
- On comparing clients with payment difficulties, working class have more clients having difficulty in payment than the clients having no difficulty in payment.
- Whereas for all the other categories, clients having payment difficulties are less than the clients with no payment difficulty.

# DISTRIBUTION OF EDUCATION TYPE

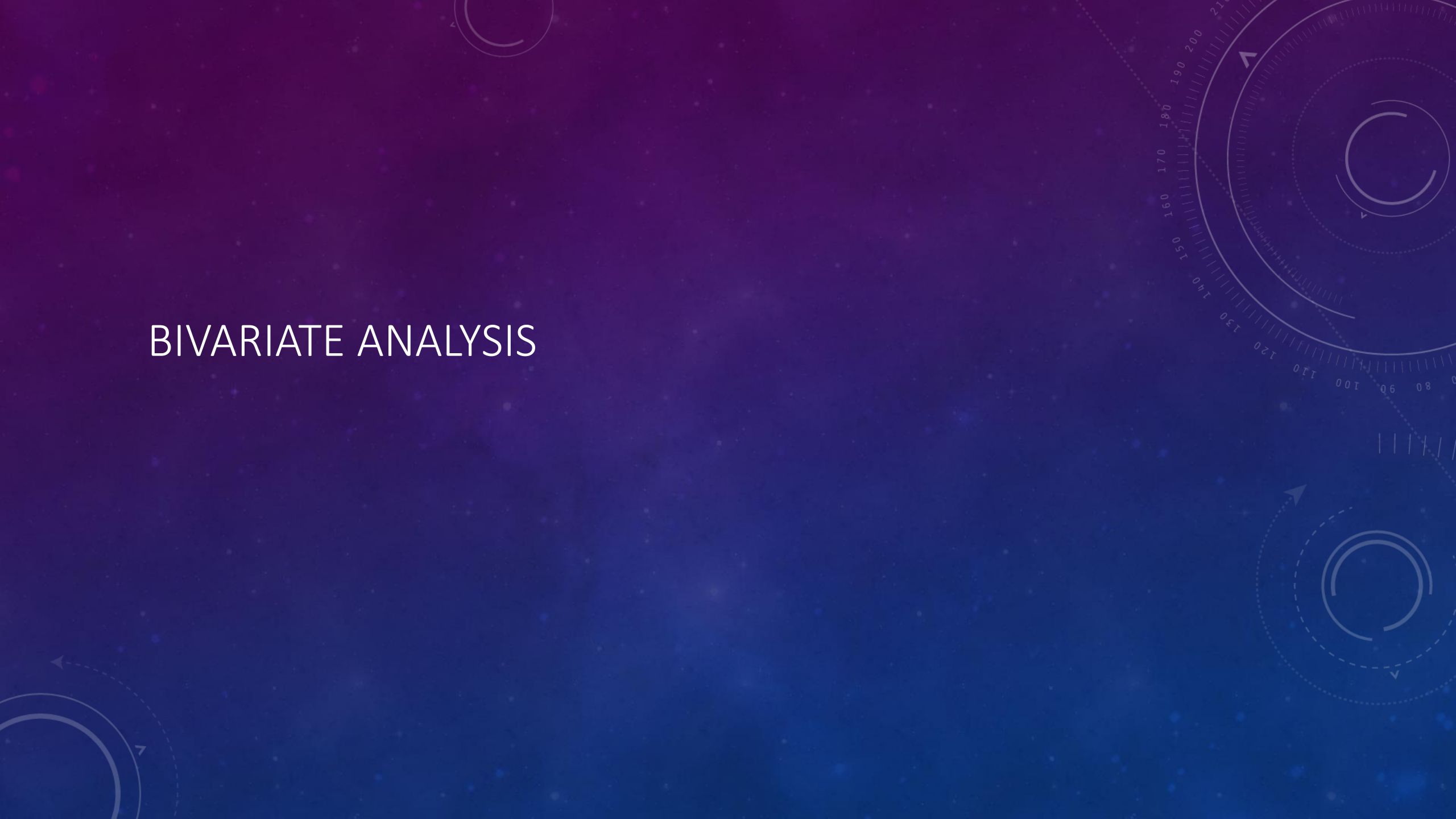


From the above plots, we can state that:

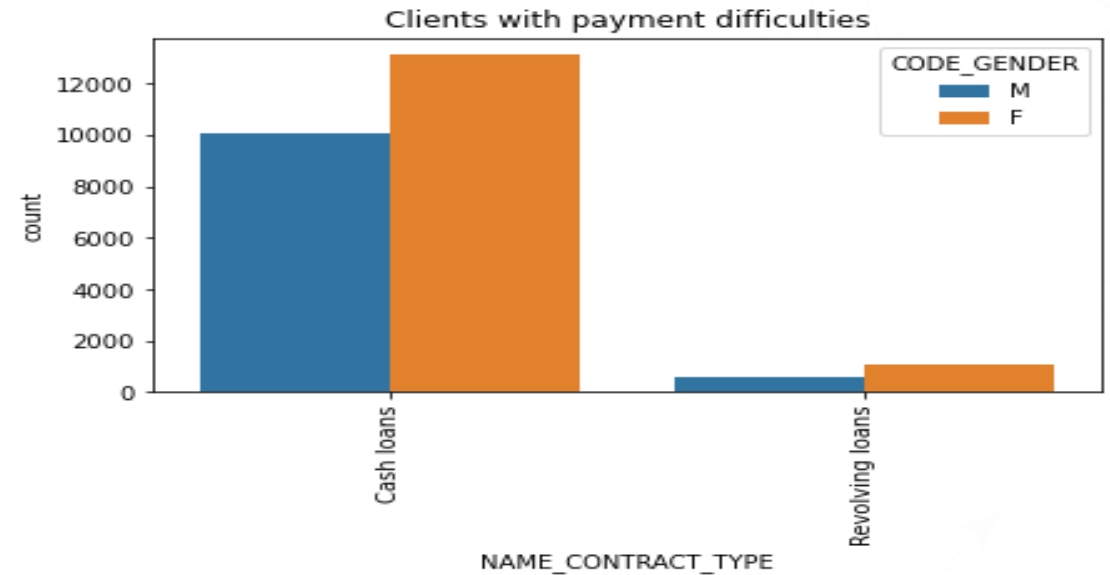
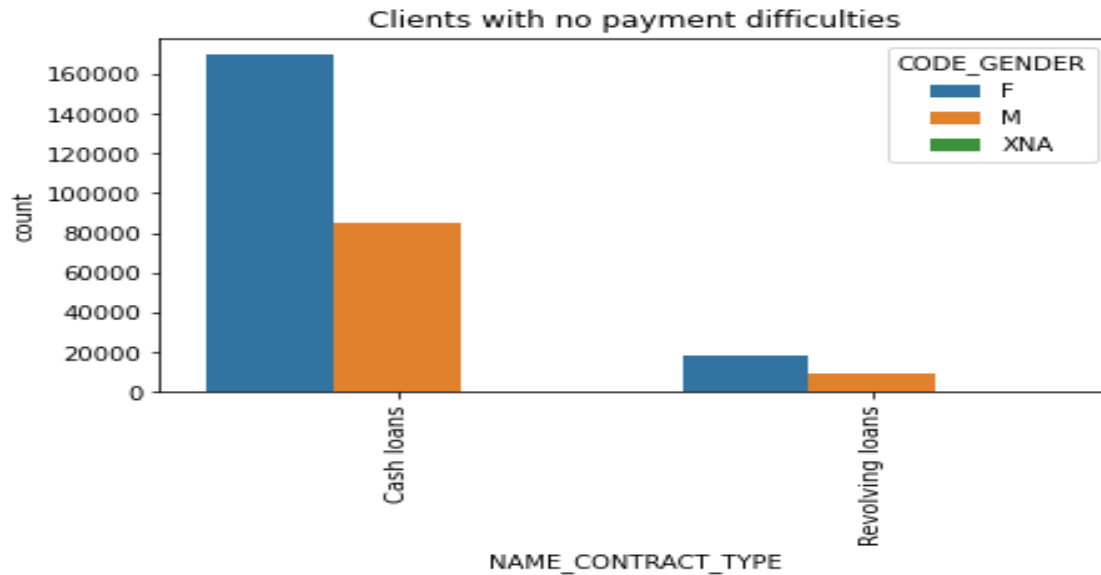
- Secondary/Secondary special educated peoples are applying for more loans than the others.
- Also these Secondary/Secondary special educated peoples have higher percentage in payment difficulties than the other.



# BIVARIATE ANALYSIS



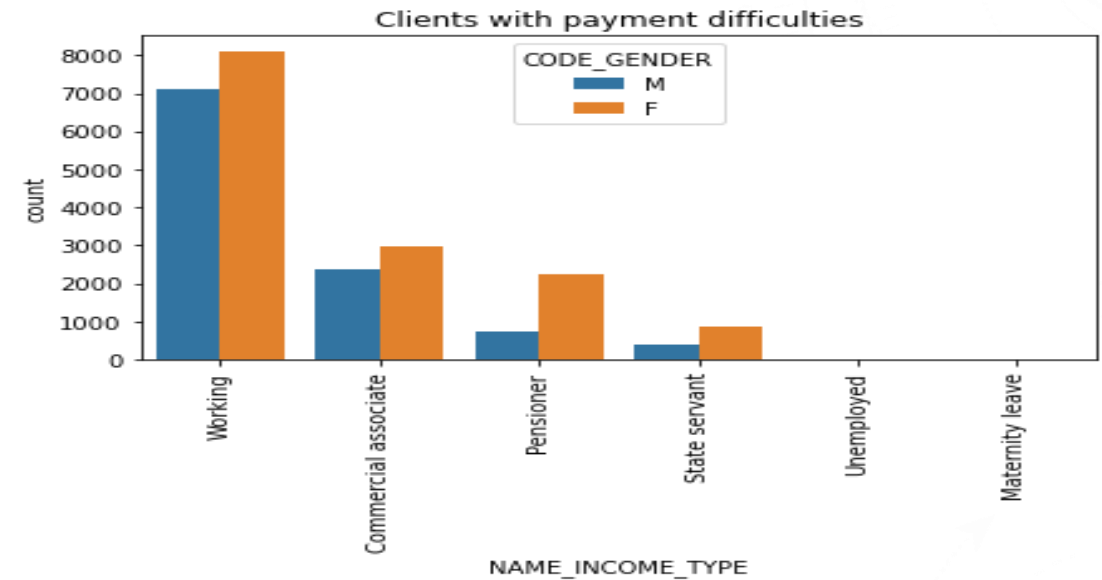
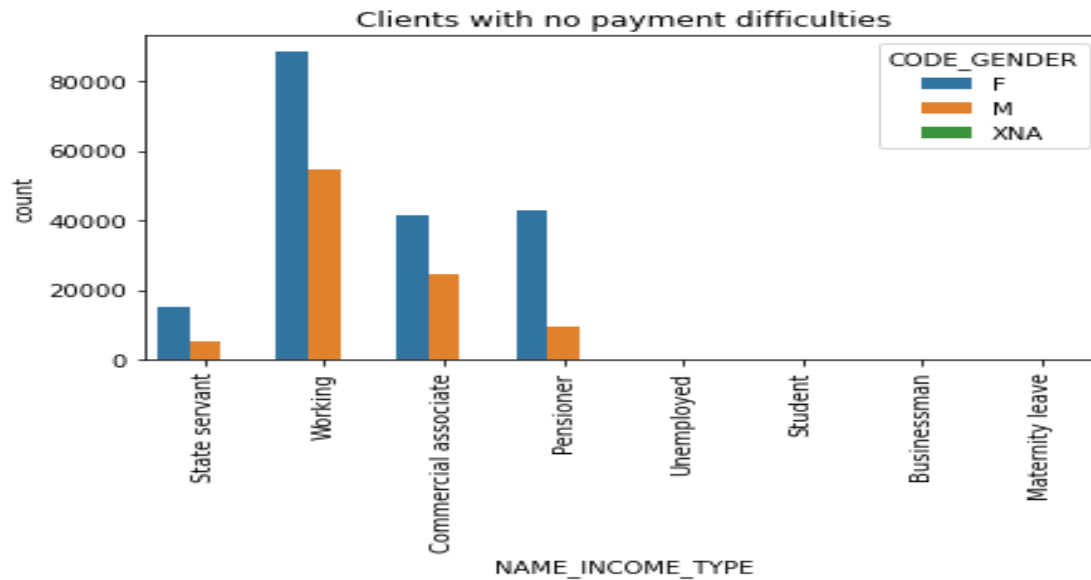
# DISTRIBUTION OF CONTRACT TYPE & GENDER



From the above plots, we can state that:

- For both type of loans, more number of females are applying for loans.
- As number of females of applying is more, more female clients having payment difficulties.

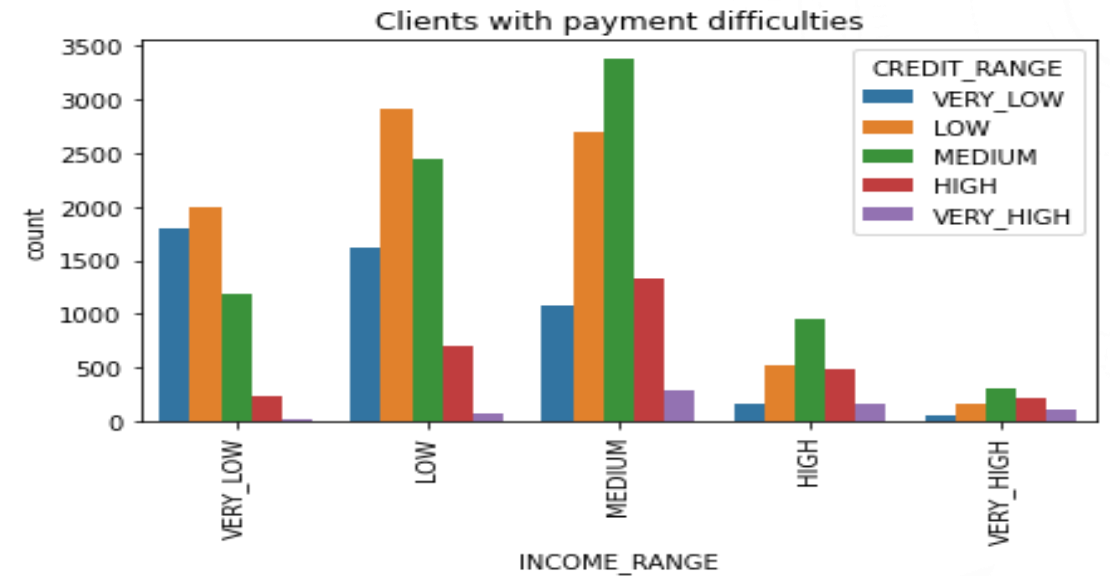
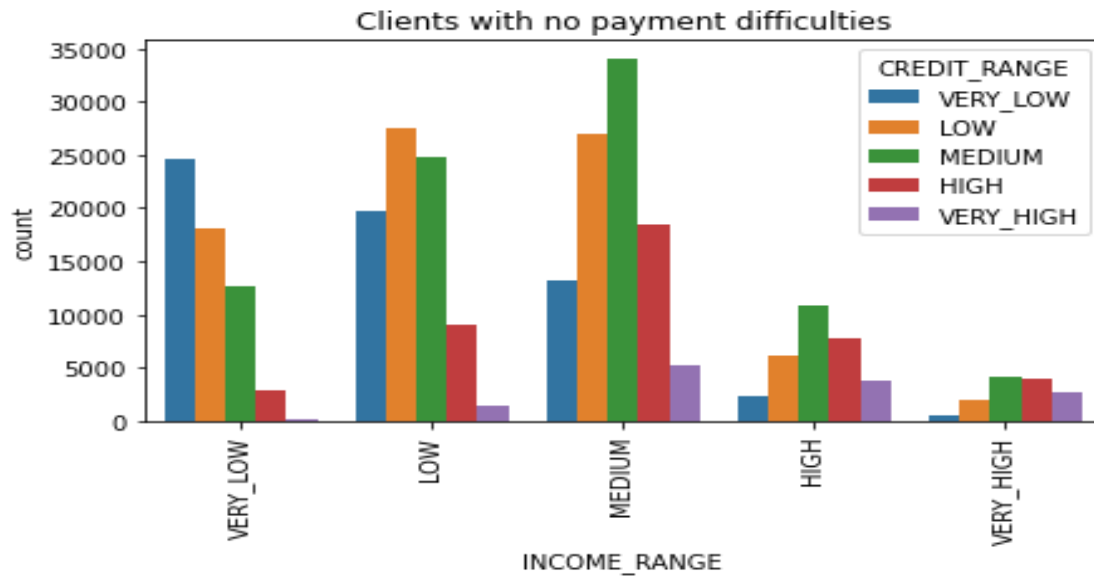
# DISTRIBUTION OF INCOME TYPE & GENDER



From the above plots, we can state that:

- Working category is applying for more loans.
- As total count for female applicants for loan is high, we can conclude by the data, less percentage of females are having difficulty in each category.

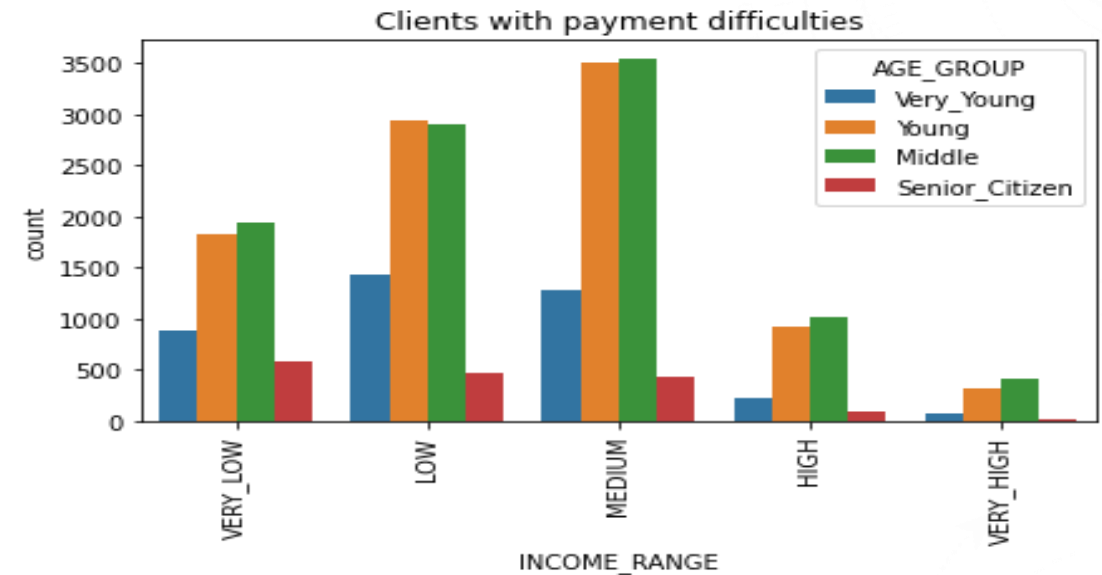
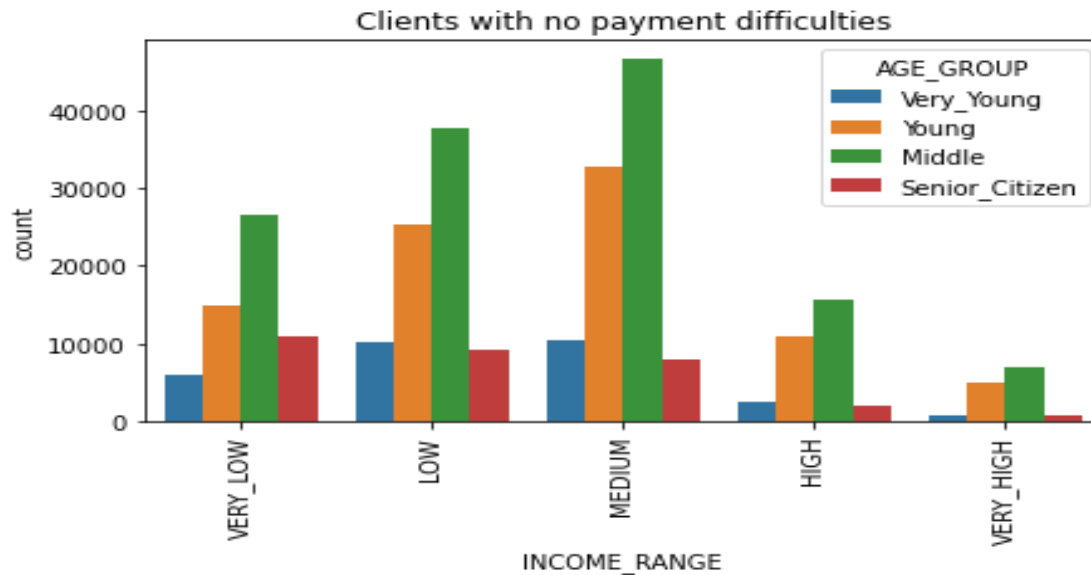
# DISTRIBUTION OF INCOME & CREDIT



From the above plots, we can state that:

- Client with medium income range are applying for more loans, whereas clients with very high income loans are not applying for much loans.
- For clients with no payment difficulties, medium income range clients are applying for medium credit range of loans, low income range are applying more for low credit range, and very low income range clients are applying for very low credit range.
- For clients with payment difficulties, trend is almost similar with clients with no payment difficulties with a slight variation in very low income range, in which clients are applying for low credit range more than very low credit range.

# DISTRIBUTION OF INCOME & AGE



From the above plots, we can state that:

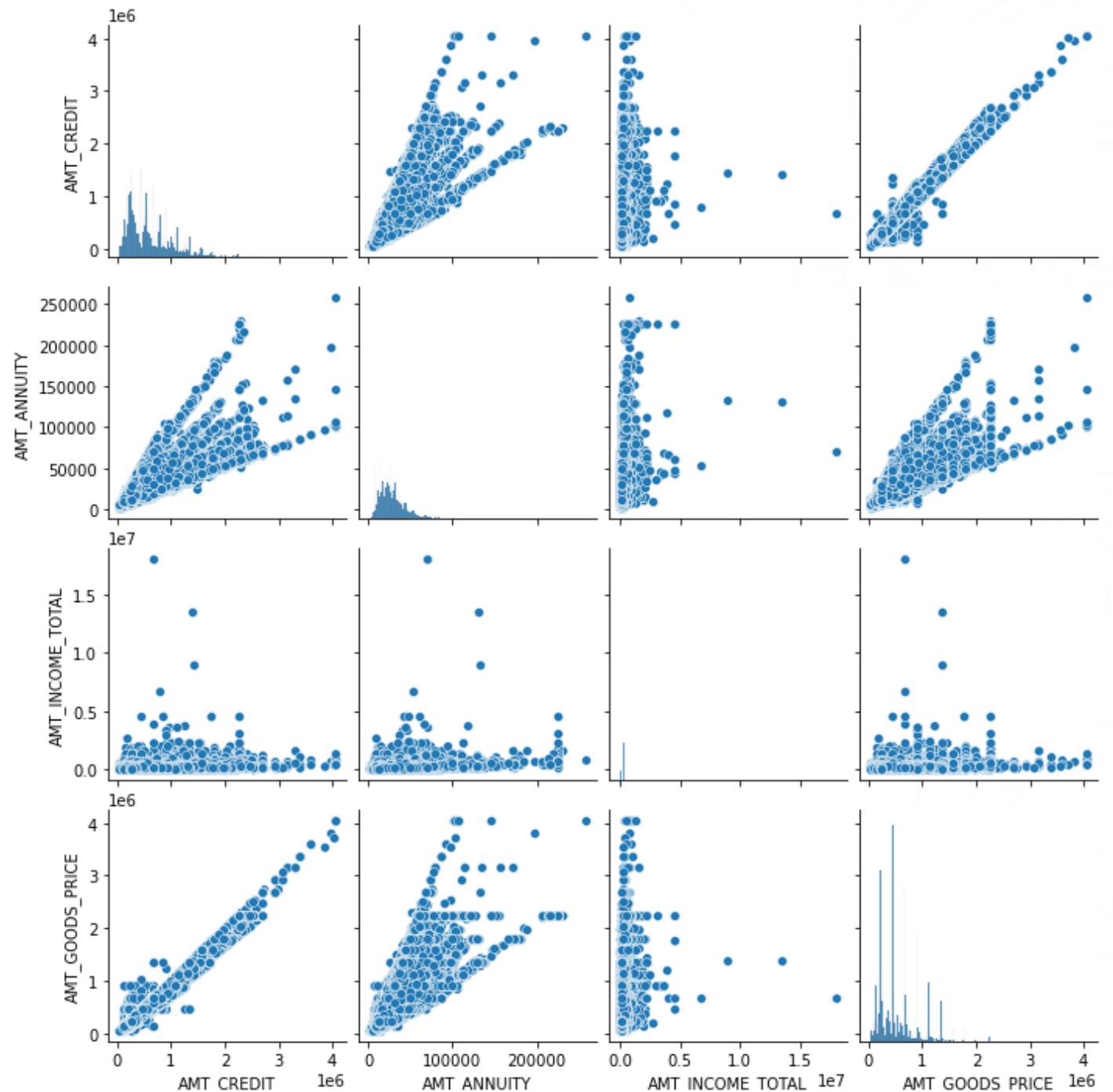
- Clients of age group young and very young have higher percentage in difficulty in payment with respect to clients with no difficulty in payment when their income is in range of very low to medium.
- As middle age group persons are also having difficulties, but on comparison their percentage of clients with payment difficulty is less. Also percentage decreases as their income group increases.
- We can also target senior citizen, their percentage for clients with payment difficulties is least, but the count of their applications is also not very high irrespective of their income group.



RELATION BETWEEN CREDIT, ANNUITY, INCOME & GOODS PRICE  
FOR CLIENTS WITH NO PAYMENT DIFFICULTY (TARGET = 0)

From the plots, we can say

- There is upward trend, high correlation between CREDIT AMOUNT and GOODS PRICE.
- There is upward trend, high correlation between ANNUITY and GOODS PRICE.
- Hence there is high correlation between ANNUITY and CREDIT AMOUNT which can be seen in the plots also.
- Density for ANNUITY, CREDIT AMOUNT and GOODS PRICE is high in lower section in plots plotted against INCOME which represents, maximum loan applications are not from high income range group.



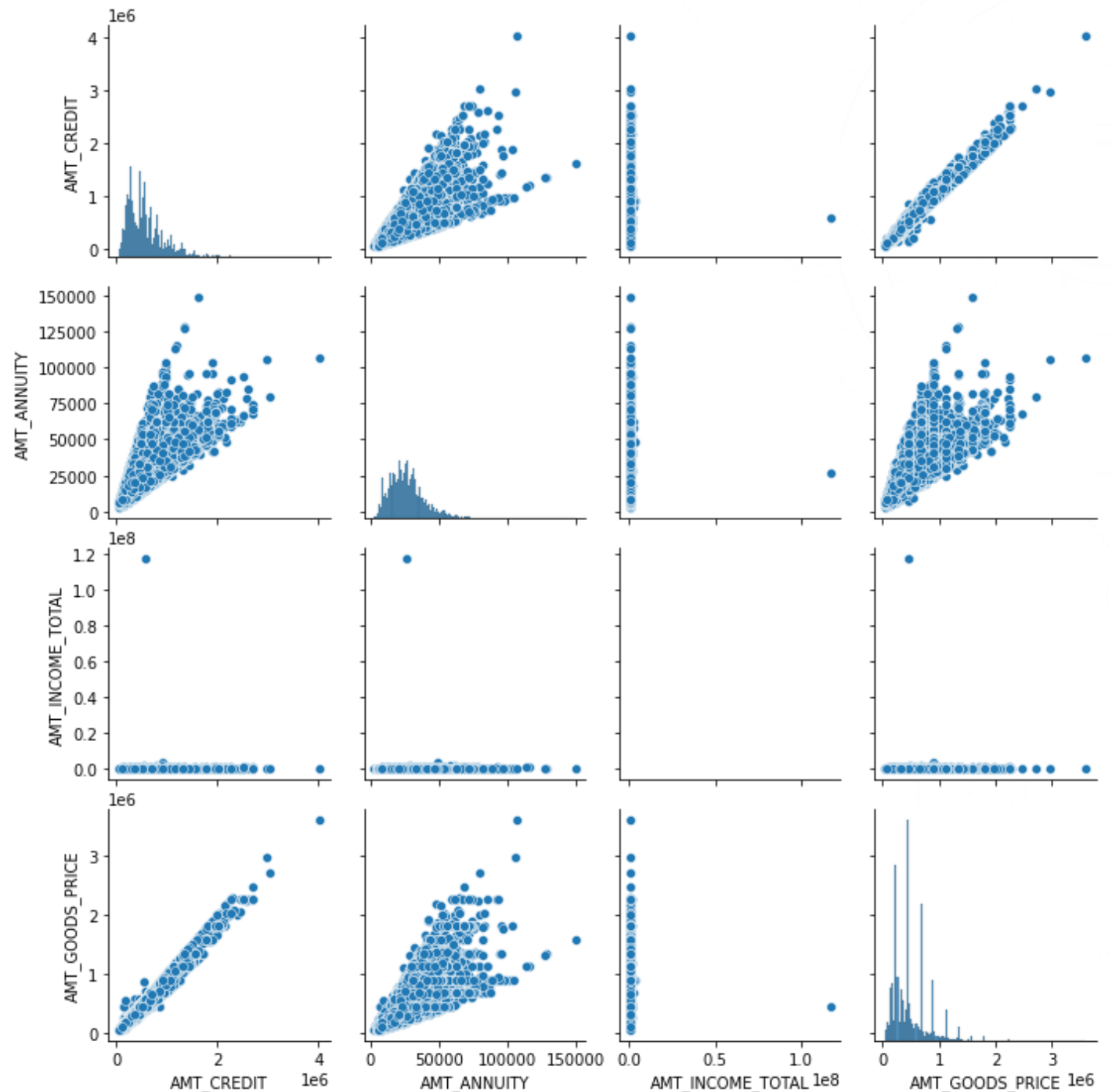


RELATION BETWEEN CREDIT, ANNUITY, INCOME & GOODS PRICE  
FOR CLIENTS WITH PAYMENT DIFFICULTIES (TARGET = 1)



From the plots, we can say

- There is upward trend, high correlation between CREDIT AMOUNT and GOODS PRICE.
- There is upward trend, high correlation between ANNUITY and GOODS PRICE.
- Hence there is high correlation between ANNUITY and CREDIT AMOUNT which can be seen in the plots also.
- Density for ANNUITY, CREDIT AMOUNT and GOODS PRICE is high in lower section in plots plotted against INCOME which represents, maximum loan applications are not from high income range group.

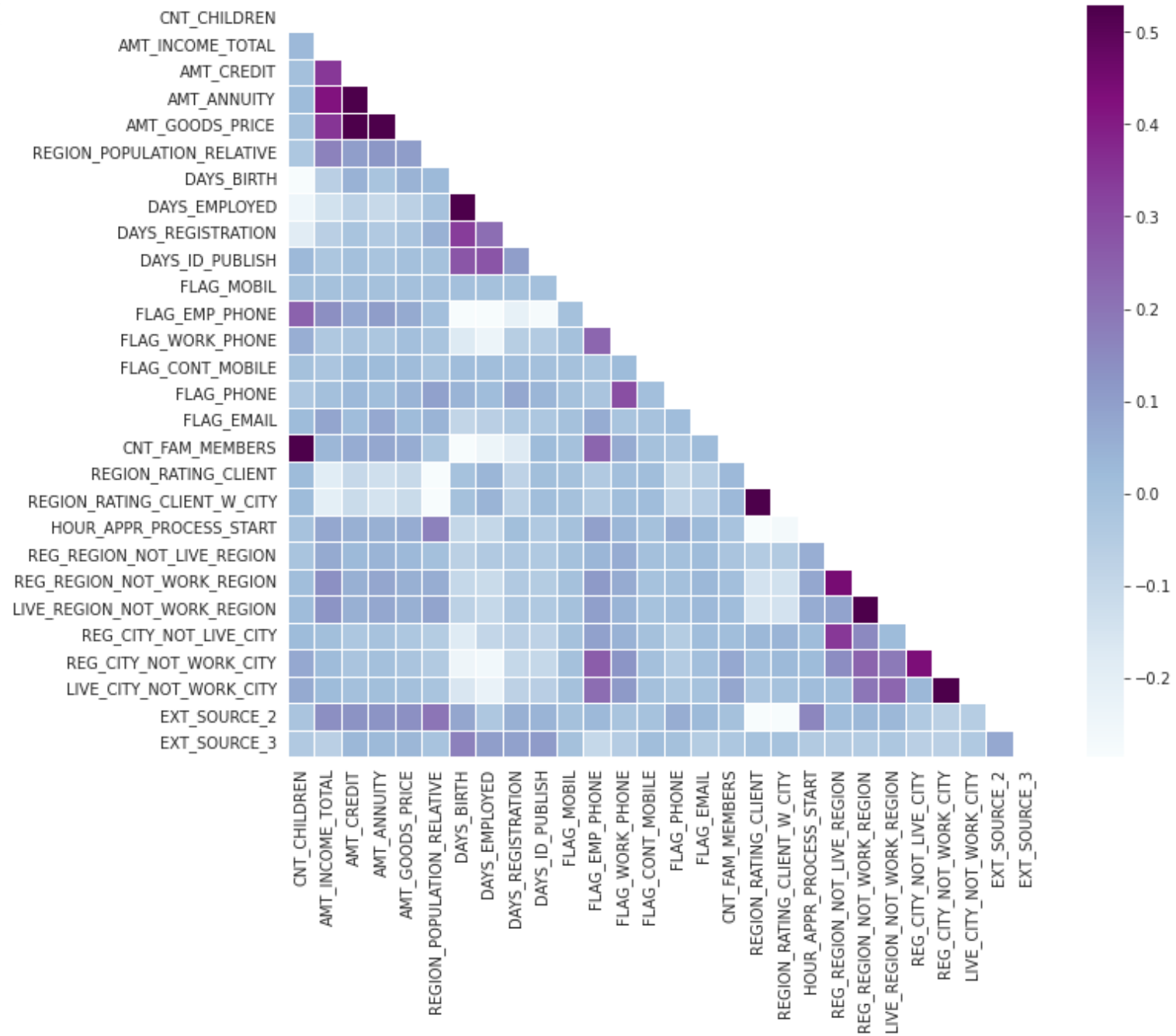


CORRELATION OF CLIENTS WITH NO PAYMENT  
DIFFICULTY (TARGET = 0)

# TOP 10 CORRELATED COLUMNS (TARGET = 0)

Column1	Column2	Column3
FLAG_EMP_PHONE	DAYS_EMPLOYED	0.999756
DAYS_EMPLOYED	FLAG_EMP_PHONE	0.999756
OBS_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	0.998510
OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.998510
AMT_CREDIT	AMT_GOODS_PRICE	0.987022
AMT_GOODS_PRICE	AMT_CREDIT	0.987022
REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY	0.950149
REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.950149
CNT_FAM_MEMBERS	CNT_CHILDREN	0.878571
CNT_CHILDREN	CNT_FAM_MEMBERS	0.878571
LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	0.861861
REG_REGION_NOT_WORK_REGION	LIVE_REGION_NOT_WORK_REGION	0.861861
DEF_30_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	0.859371
DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.859371
LIVE_CITY_NOT_WORK_CITY	REG_CITY_NOT_WORK_CITY	0.830381
REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY	0.830381
AMT_GOODS_PRICE	AMT_ANNUITY	0.776421
AMT_ANNUITY	AMT_GOODS_PRICE	0.776421
AMT_ANNUITY	AMT_CREDIT	0.771297
AMT_CREDIT	AMT_ANNUITY	0.771297

# HEATMAP FOR CLIENT WITH NO PAYMENT DIFFICULTY (TARGET = 0)



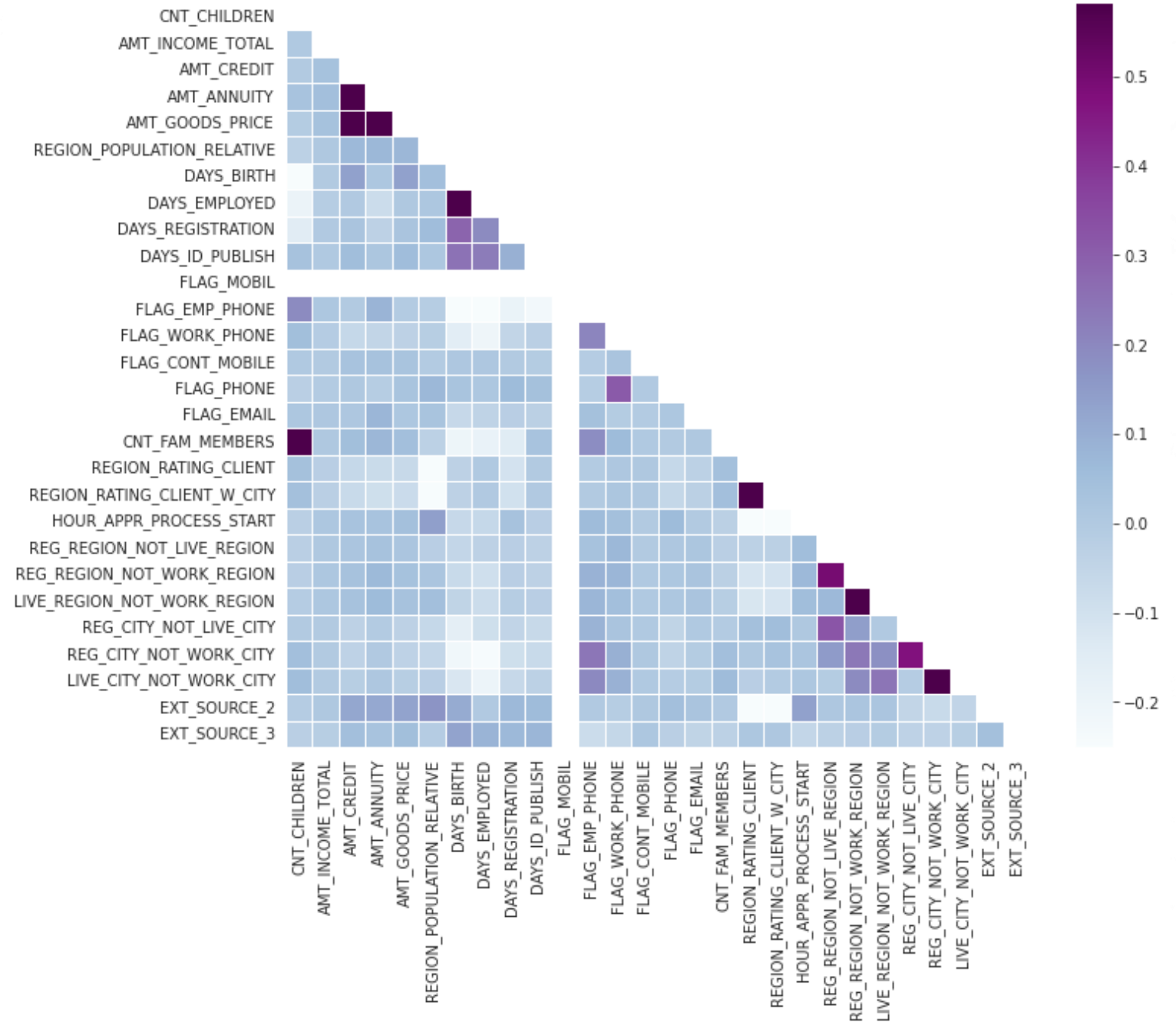
# CORRELATION OF CLIENTS WITH PAYMENT DIFFICULTIES (TARGET = 1)



# TOP 10 CORRELATED COLUMNS (TARGET = 1)

Column1	Column2	Column3
FLAG_EMP_PHONE	DAYS_EMPLOYED	0.999756
DAYS_EMPLOYED	FLAG_EMP_PHONE	0.999756
OBS_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	0.998510
OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.998510
AMT_CREDIT	AMT_GOODS_PRICE	0.987022
AMT_GOODS_PRICE	AMT_CREDIT	0.987022
REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY	0.950149
REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.950149
CNT_FAM_MEMBERS	CNT_CHILDREN	0.878571
CNT_CHILDREN	CNT_FAM_MEMBERS	0.878571
LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	0.861861
REG_REGION_NOT_WORK_REGION	LIVE_REGION_NOT_WORK_REGION	0.861861
DEF_30_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	0.859371
DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.859371
LIVE_CITY_NOT_WORK_CITY	REG_CITY_NOT_WORK_CITY	0.830381
REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY	0.830381
AMT_GOODS_PRICE	AMT_ANNUITY	0.776421
AMT_ANNUITY	AMT_GOODS_PRICE	0.776421
AMT_ANNUITY	AMT_CREDIT	0.771297
AMT_CREDIT	AMT_ANNUITY	0.771297

# HEATMAP FOR CLIENT WITH PAYMENT DIFFICULTY (TARGET = 1)





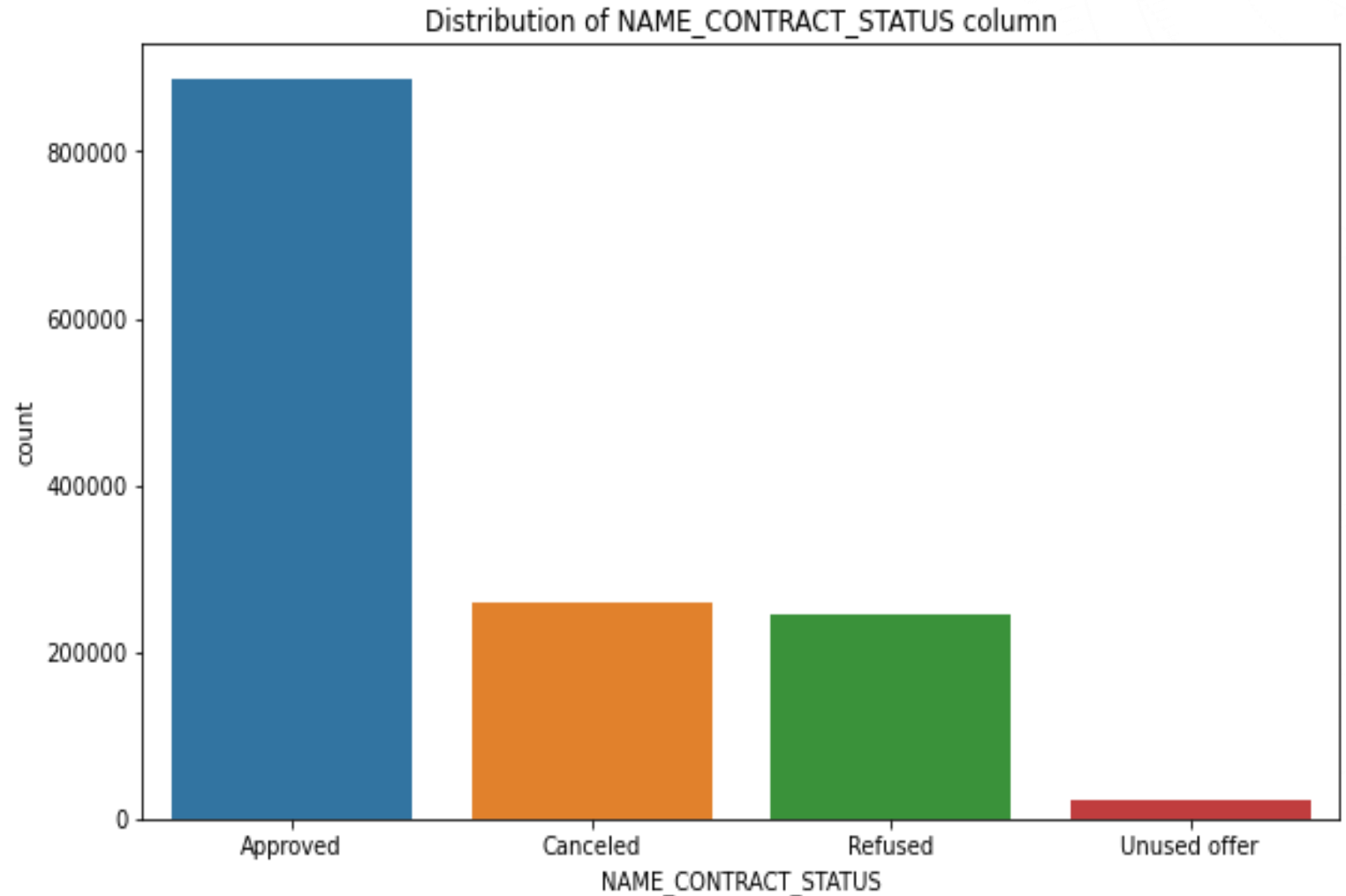
# ANALYSIS AFTER MERGING PREVIOUS AND CURRENT DATA



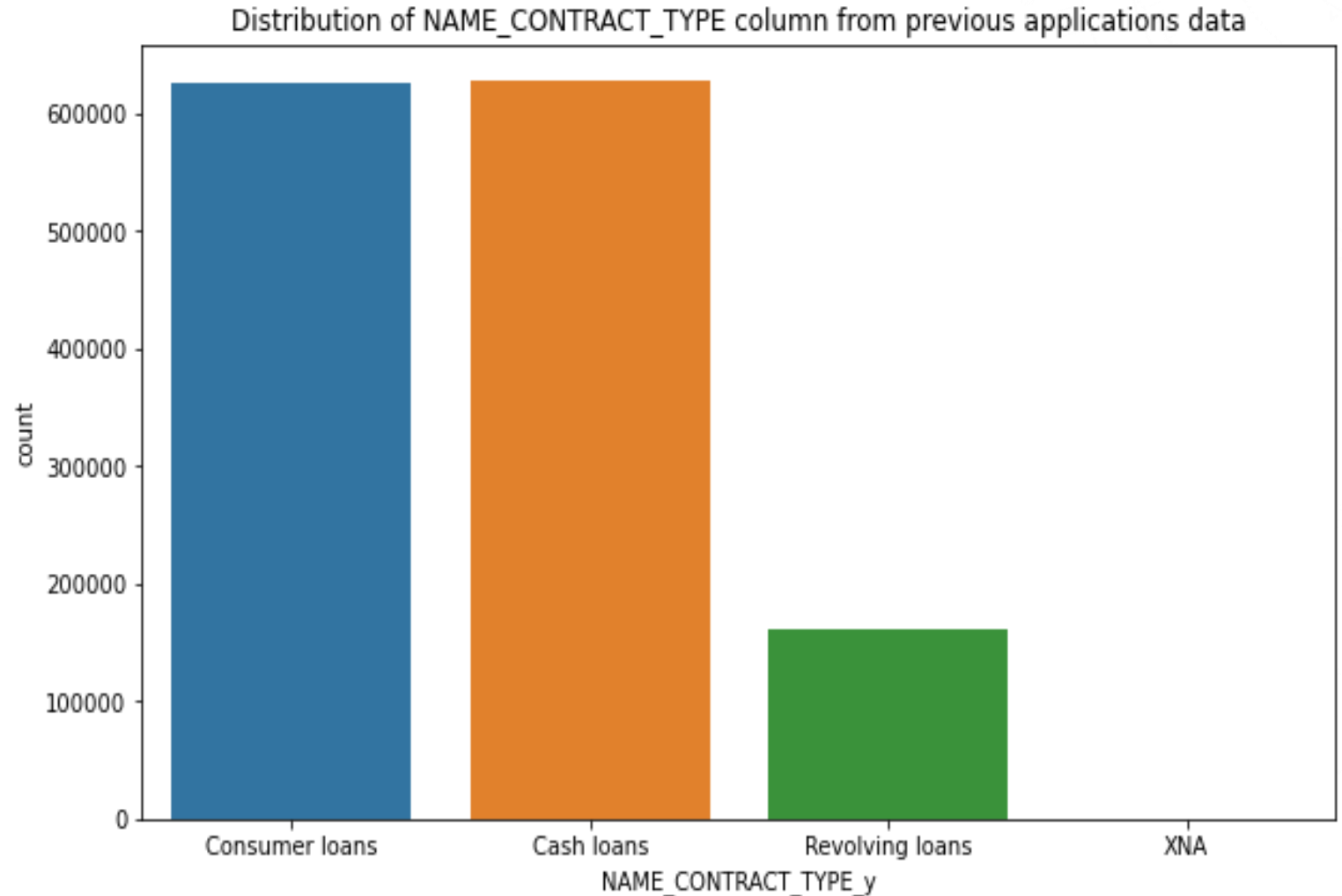
# DISTRIBUTION OF CONTRACT STATUS

From the plot, we can state that:

- Very few applications are in unused offer category, and most of the applications are in approved category



# DISTRIBUTION OF CONTRACT TYPE FROM PREVIOUS DATA



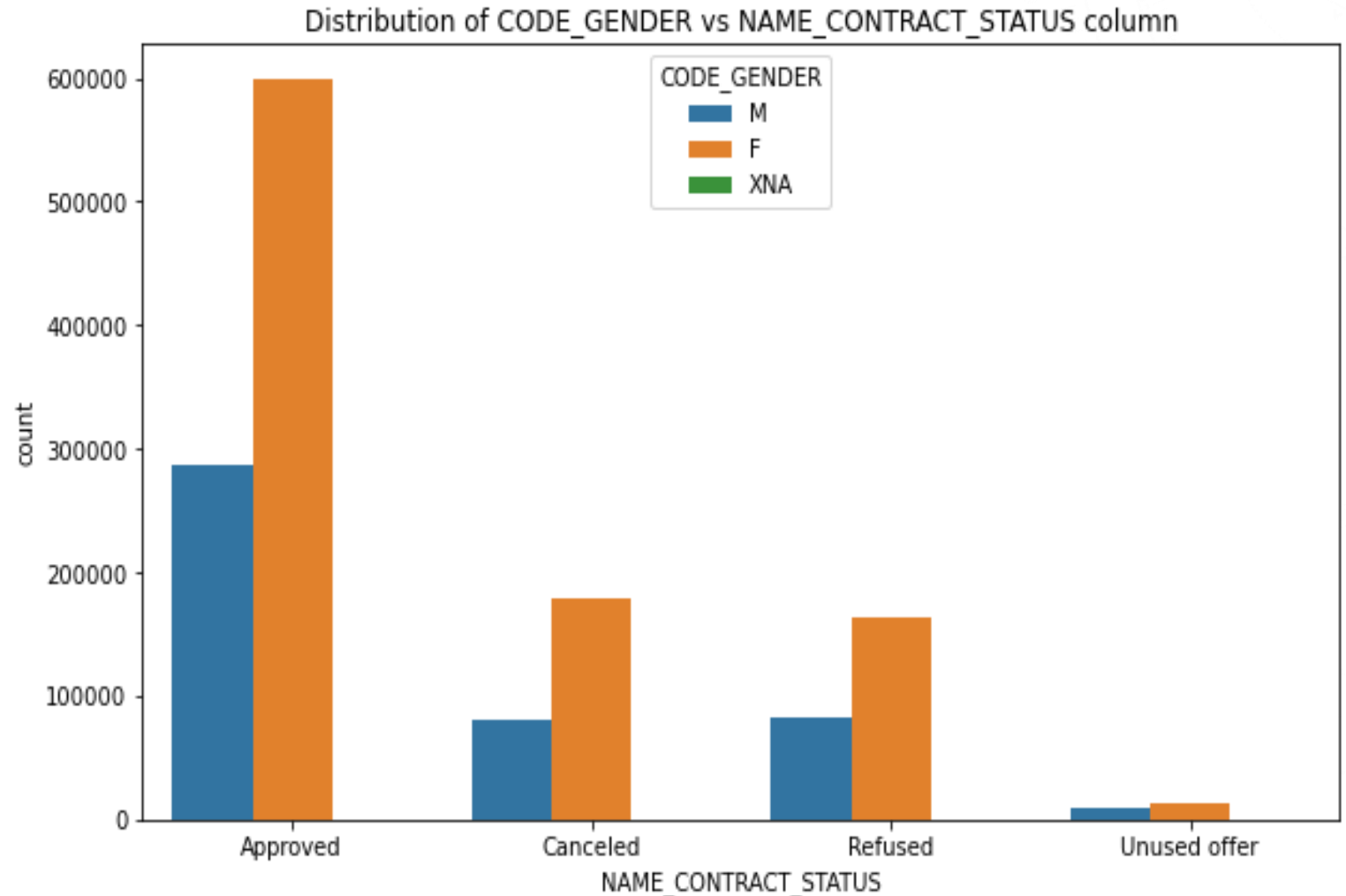
From the plot, we can state that:

- Very few applications are of Revolving types, and most of the applications are for either Cash loans or consumer loans

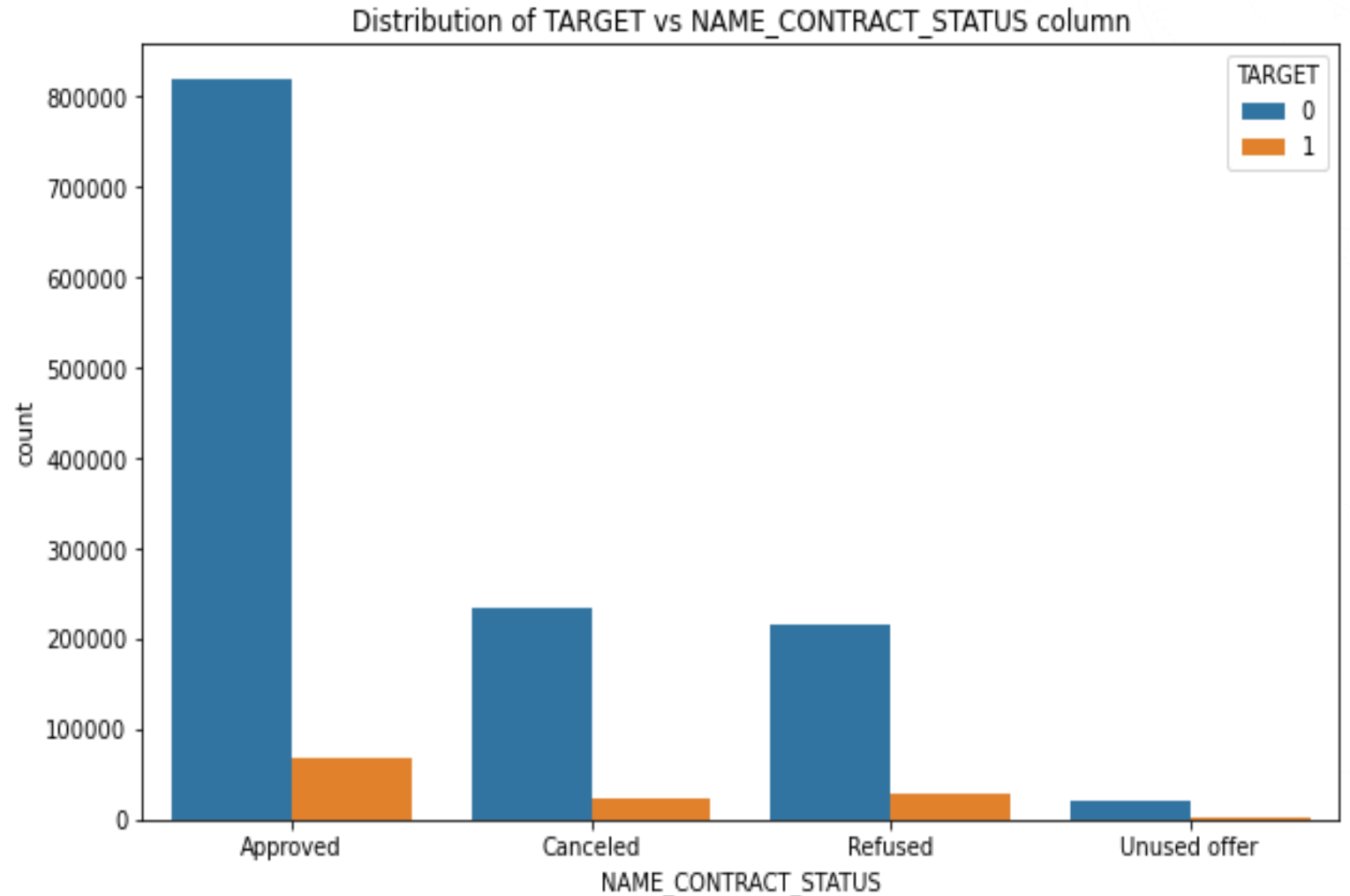
# DISTRIBUTION OF CONTRACT STATUS AND GENDER

From the plot, we can state that:

- Most of the loans are applied by females, hence number of females are high in this plot also in each category



# DISTRIBUTION OF CONTRACT STATUS AND TARGET



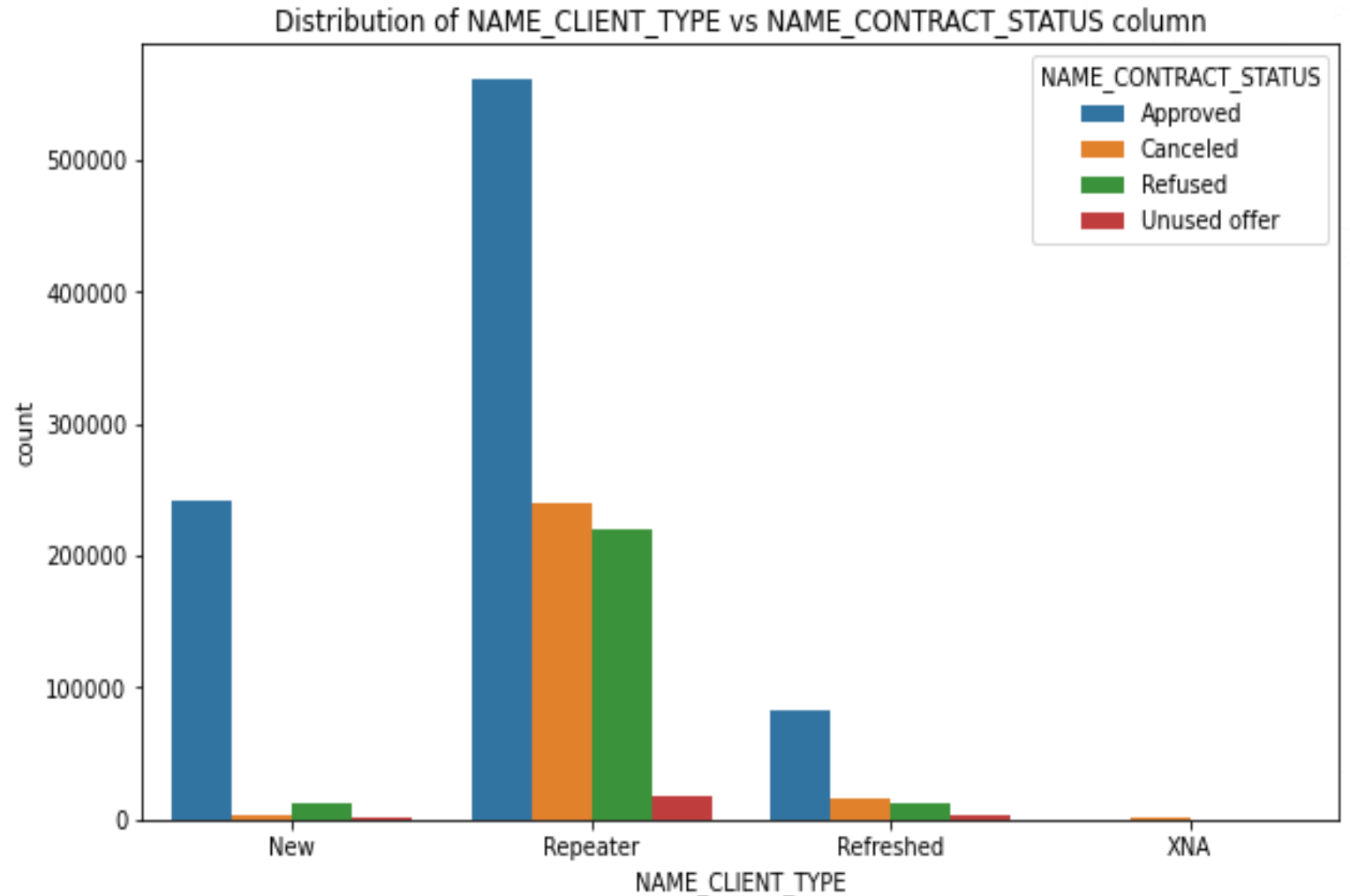
From the plot, we can state that:

- Very few percentage of loans are approved to the clients who have difficulties in payment in comparison to clients having no payment difficulties.

# DISTRIBUTION OF CONTRACT STATUS AND CLIENT TYPE

From the plot, we can state that:

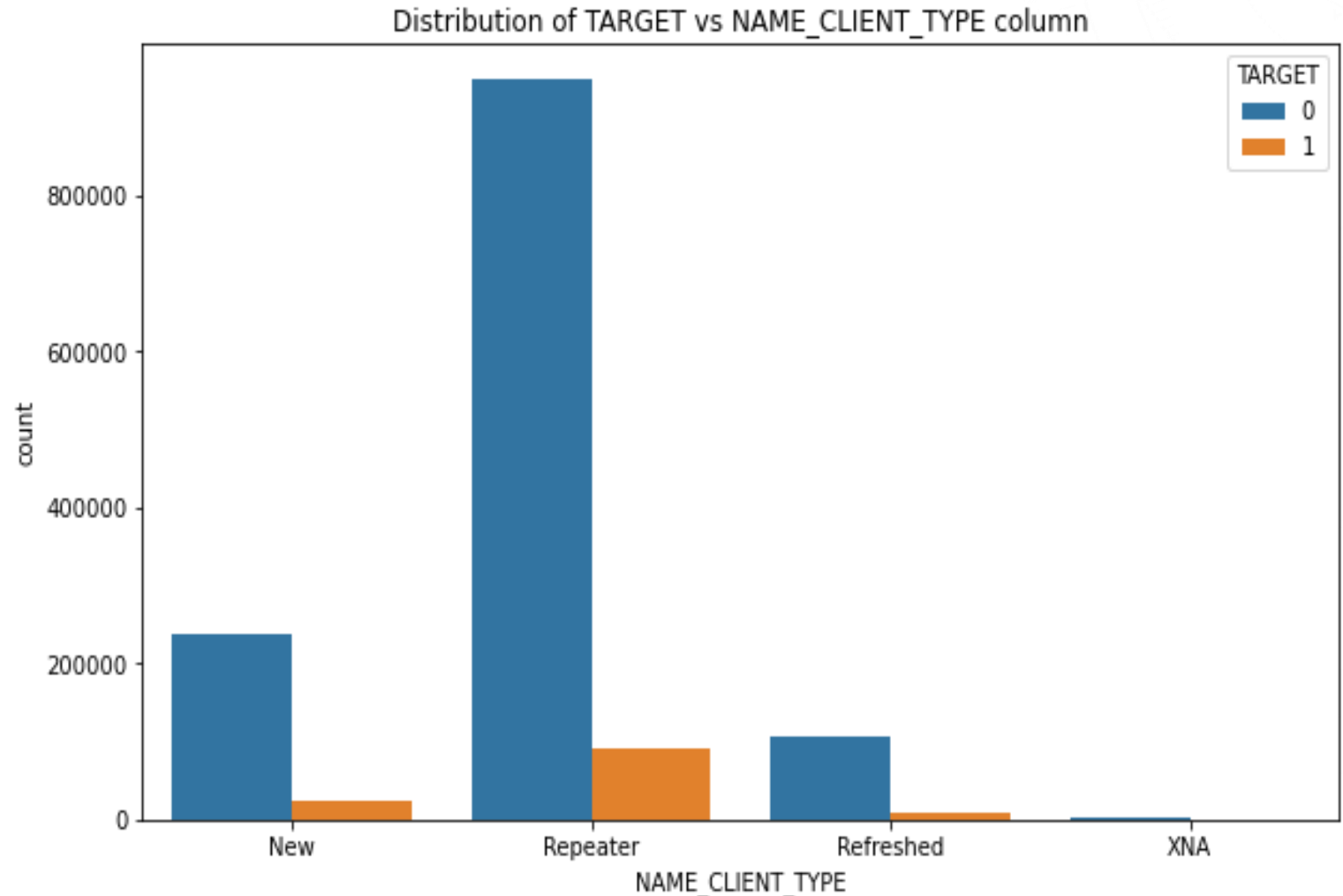
- Repeater type clients are getting max number of loans approved, but they also have max number of loans in other status.
- New clients are also getting loans approved, much more than refused or cancelled.



# DISTRIBUTION OF CONTRACT STATUS AND CLIENT TYPE

From the plot, we can state that:

- Maximum numbers of clients having payment difficulties is from repeater clients only. But they also have maximum number of clients having no difficulties.
- Maximum new clients have no payment difficulties
- Analyzing by percentage, repeater clients have very low percentage of the clients having difficulty in payments.



# CONCLUSION

Banks should focus more on below categories:

- INCOME TYPE: Pensioners, State Servants
- EDUCATION: Higher Education
- AGE GROUP: Middle Aged, Senior Citizens
- CLIENT TYPE: Repeater Clients

THANK YOU